Coordinate Descent

Nipun Batra

IIT Gandhinagar

August 1, 2025

Coordinate Descent for Unregularized Regression

- Express error as a difference of y_i and $\hat{y}_i$

$$\hat{y}_i = \sum_{j=0}^{d} \theta_j x_i^j = \theta_0 x_i^0 + \theta_1 x_i^1 + \theta_2 x_i^2 \ldots\ldots + \theta_d x_i^d \qquad (1)$$

$$\epsilon_i = y_i - \hat{y}_i \qquad (2)$$

$$= y_i - \theta_0 x_i^0 + \theta_1 x_i^1 + \ldots\ldots + \theta_d x_i^d \qquad (3)$$

$$= y_i - \sum_{j=0}^{d} \theta_j x_i^j \qquad (4)$$

Coordinate Descent for Unregularized regression

$$\sum_{i=1}^{N} \epsilon^2 = RSS = \sum_{i=1}^{N} \left(y_i - \left(\theta_0 x_i^0 + \ldots \quad \theta_j x_i^j + \theta_d x_i^d \right) \right)^2$$

Coordinate Descent for Unregularized regression

$$\sum_{i=1}^{N} \epsilon^2 = RSS = \sum_{i=1}^{N} \left(y_i - \left(\theta_0 x_i^0 + \ldots \quad \theta_j x_i^j + \theta_d x_i^d \right) \right)^2$$

$$\frac{\partial \, \mathrm{RSS} \, (\theta_j)}{\partial \theta_j} = 2 \sum_{i=1}^{N} \left(y_i - \left(\theta_0 x_i^0 + \ldots \quad \theta_j x_i^j + \ldots \right) \right) \left(-x_i^j \right)$$

Coordinate Descent for Unregularized regression

$$\sum_{i=1}^{N} \epsilon^2 = RSS = \sum_{i=1}^{N} \left(y_i - \left(\theta_0 x_i^0 + \ldots \quad \theta_j x_i^j + \theta_d x_i^d \right) \right)^2$$

$$\frac{\partial \, \mathrm{RSS} \, (\theta_j)}{\partial \theta_j} = 2 \sum_{i=1}^{N} \left(y_i - \left(\theta_0 x_i^0 + \ldots \quad \theta_j x_i^j + \ldots \right) \right) \left(-x_i^j \right)$$

$$= 2 \sum_{i=1}^{N} \left(y_i - \left(\theta_0 x_i^0 + \ldots + \theta_d x_i^d \right) \right) \left(-x_i^j \right) + 2 \sum_{i=1}^{N} \theta_j (x_i^j)^2$$

Coordinate Descent for Unregularized regression

$$\sum_{i=1}^{N} \epsilon^2 = RSS = \sum_{i=1}^{N} \left(y_i - \left(\theta_0 x_i^0 + \ldots \quad \theta_j x_i^j + \theta_d x_i^d \right) \right)^2$$

$$\frac{\partial \operatorname{RSS}(\theta_j)}{\partial \theta_j} = 2 \sum_{i=1}^{N} \left(y_i - \left(\theta_0 x_i^0 + \ldots \quad \theta_j x_i^j + \ldots \right) \right) \left(-x_i^j \right)$$

$$= 2 \sum_{i=1}^{N} \left(y_i - \left(\theta_0 x_i^0 + \ldots + \theta_d x_i^d \right) \right) \left(-x_i^j \right) + 2 \sum_{i=1}^{N} \theta_j (x_i^j)^2$$

where:

$$\hat{y}_i^{(-j)} = \theta_0 x_i^0 + \ldots + \theta_d x_i^d$$

is $\hat{y}_i$ without θ_j

Coordinate Descent for Unregularized regression

$$Set \frac{\partial \operatorname{RSS}(\theta_j)}{\partial \theta_j} = 0$$

$$\theta_j = \sum_{i=1}^{N} \frac{\left(y_i - \left(\theta_0 x_i^0 + \ldots + \ldots + \theta_d x_i^d\right)\right)\left(x_i^j\right)}{\left(x_i^j\right)^2} = \frac{\rho_j}{z_j}$$

$$\rho_j = \sum_{i=1}^{N} x_i^j \left(y_i - \hat{y}_i^{(-j)}\right)$$

$$z_j = \sum_{i=1}^{N} \left(x_i^j\right)^2$$

z_j is the squared of ℓ_2 norm of the jth feature

UNDERSTANDING ρ_j IN COORDINATE DESCENT

$$\rho_j = \sum_{i=1}^{N} x_i^j \left(y_i - \hat{y}_i^{(-j)} \right)$$

UNDERSTANDING β_j IN COORDINATE DESCENT

$$\beta_j = \sum_{i=1}^{N} x_i^j \left(y_i - \hat{y}_i^{(-j)} \right)$$
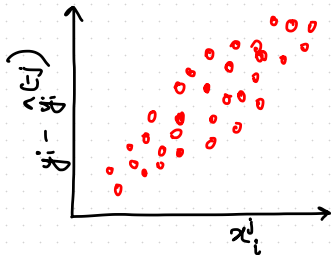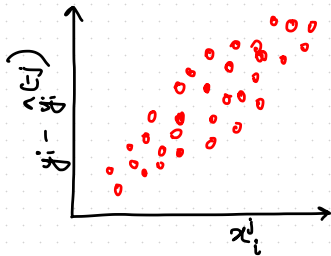
CASE 1

x_i^j STRONG +VE CORR. WITH $y_i - \hat{y}_i^{(-j)}$

UNDERSTANDING ρ_j IN COORDINATE DESCENT

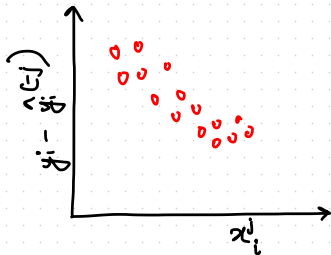$$\rho_j = \sum_{i=1}^{N} x_i^j \left(y_i - \hat{y}_i^{(-j)} \right)$$

CASE 1

x_i^j STRONG +VE CORR.
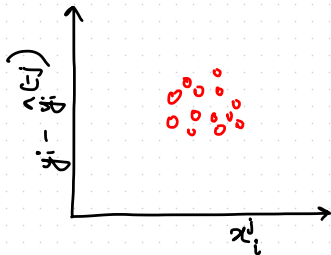WITH $y_i - \hat{y}_i^{(-j)}$

$$\Downarrow$$

j^{th} FEATURE IS IMPT.
AND ITS COEFFICIENT
+VE

UNDERSTANDING β_j IN COORDINATE DESCENT

$$\beta_j = \sum_{i=1}^{N} x_i^j \left(y_i - \hat{y}_i^{(-j)} \right)$$

CASE II

x_i^j STRONG $-VE$ CORR.

WITH $y_i - \hat{y}_i^{(-j)}$

$\Downarrow$

j^{th} FEATURE IS IMPT.

AND ITS COEF. $-VE$

UNDERSTANDING β_j IN COORDINATE DESCENT

$$\beta_j = \sum_{i=1}^{N} x_i^j \left(y_i - \hat{y}_i^{(-j)} \right)$$

CASE II

x_i^j WEAK CORR.

WITH $y_i - \hat{y}_i^{(-j)}$

$\Downarrow$

j^{th} FEATURE IS **NOT** IMPT.

AND ITS COEF. $\to 0$

Coordinate Descent for Lasso Regression

$$\text{Minimize} \underbrace{\sum_{i=1}^{N} \epsilon^2 + \delta^2 \left\{ |\theta_0| + |\theta_1| + \ldots |\theta_j| + \ldots |\theta_d| \right\}}_{\textit{LASSO OBJECTIVE}}$$

$$\frac{\partial}{\partial \theta_j}(\text{LASSO OBJECTIVE}) = -2\rho_j + 2\theta_j z_j + \delta^2 \frac{\partial}{\partial \theta_j} |\theta_j|$$

$$\frac{\partial}{\partial \theta_j} |\theta_j| = \left\{ \begin{array}{cc} 1 & \theta_j > 0 \\ [-1, 1] & \theta_j = 0 \\ -1 & \theta_j < 0 \end{array} \right.$$

Coordinate Descent for Lasso Regression

- **Case 1:** $\theta_j > 0$

$$2\rho_j + 2\theta_j z_j + \delta^2 = 0$$

$$\theta_j = \frac{\rho_j - \frac{\delta^2}{2}}{z_j}$$

$$\rho_j > \frac{\delta^2}{2} \Rightarrow \theta_j = \frac{\rho_j - \frac{\delta^2}{2}}{z_j}$$

Coordinate Descent for Lasso Regression

- **Case 1:** $\theta_j > 0$

$$2\rho_j + 2\theta_j z_j + \delta^2 = 0$$

$$\theta_j = \frac{\rho_j - \frac{\delta^2}{2}}{z_j}$$

$$\rho_j > \frac{\delta^2}{2} \Rightarrow \theta_j = \frac{\rho_j - \frac{\delta^2}{2}}{z_j}$$

- **Case 2:** $\theta_j < 0$

$$\rho_j < \frac{\delta^2}{2} \Rightarrow \theta_j = \frac{\rho_j + \delta^2/2}{z_j} \tag{5}$$

Coordinate Descent for Lasso Regression

- **Case 3:** $\theta_j = 0$

$$\frac{\partial}{\partial \theta_j}(\text{LASSO OBJECTIVE}) = -2\rho_j + 2\theta_j z_j + \delta^2 \underbrace{\frac{\partial}{\partial \theta_j}|\theta_j|}_{[\text{-1,1}]}$$

$$\underbrace{\epsilon\left[-2\rho_j - \delta^2, -2\rho_j + \delta^2\right]}_{\{0\} \text{ lies in this range}}$$

$$-2\rho_j - \delta^2 \leq 0 \text{ and } -2\rho_j - \delta^2 \leq 0$$

$$-\frac{\delta^2}{2} \leq \rho_j \leq \frac{\delta^2}{2} \Rightarrow \theta_j = 0$$
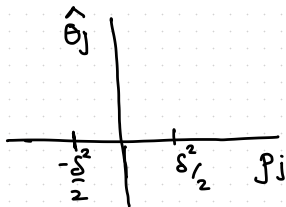
Summary of Lasso Regression

$$
\theta_j = \left[
\begin{array}{ccc}
\frac{\rho_j + \frac{\delta^2}{2}}{z_j} & \textit{if} & \rho_j < -\frac{\delta^2}{2} \\
0 & \textit{if} & -\frac{\delta^2}{2} \leq \rho_j \leq \frac{\delta^2}{2} \\
\frac{\rho_j - \frac{\delta^2}{2}}{z_j} & \textit{if} & \rho_j > \frac{\delta^2}{2}
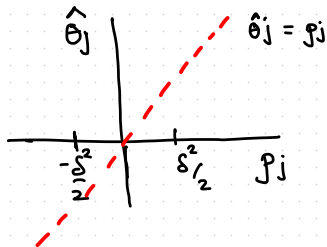\end{array}
\right]
\qquad (6)
$$

LASSO (SOFT) THRESHOLDING

$$\theta_j = \begin{cases} \dfrac{\rho_j + \delta^2/2}{z_j} & \text{if} \quad \rho_j < -\delta^2/2 \\[2ex] 0 & \text{if} \quad -\dfrac{\delta^2}{2} \leq \rho_j \leq \dfrac{\delta^2}{2} \\[2ex] \dfrac{\rho_j - \delta^2/2}{z_j} & \text{if} \quad \rho_j > \delta^2/2 \end{cases}$$
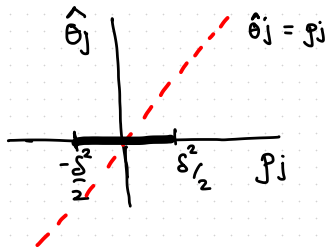
LASSO (SOFT) THRESHOLDING

$$\hat{\theta}_j = \begin{cases} \dfrac{\rho_j + \delta^2/2}{z_j} & \text{if} \quad \rho_j < -\delta^2/2 \\[2mm] 0 & \text{if} \quad -\dfrac{\delta^2}{2} \leq \rho_j \leq \dfrac{\delta^2}{2} \\[2mm] \dfrac{\rho_j - \delta^2/2}{z_j} & \text{if} \quad \rho_j > \delta^2/2 \end{cases}$$
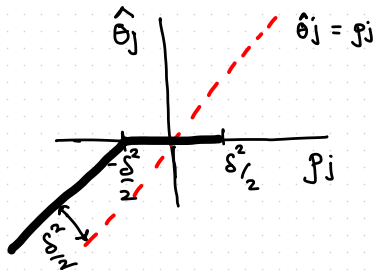
LASSO (SOFT) THRESHOLDING

$$\hat{\theta}_j = \begin{cases} \dfrac{\rho_j + \delta^2/2}{z_j} & \text{if} \quad \rho_j < -\delta^2/2 \\[2mm] 0 & \text{if} \quad -\dfrac{\delta^2}{2} \leq \rho_j \leq \dfrac{\delta^2}{2} \\[2mm] \dfrac{\rho_j - \delta^2/2}{z_j} & \text{if} \quad \rho_j > \delta^2/2 \end{cases}$$
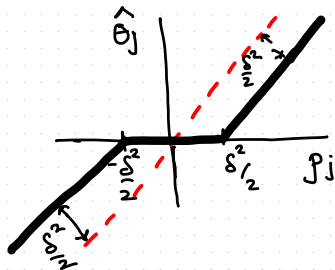
LASSO (SOFT) THRESHOLDING

$$\hat{\theta}_j = \begin{cases} \dfrac{\rho_j + \delta^2/2}{z_j} & \text{if} \quad \rho_j < -\delta^2/2 \\[2mm] 0 & \text{if} \quad -\dfrac{\delta^2}{2} \leq \rho_j \leq \dfrac{\delta^2}{2} \\[2mm] \dfrac{\rho_j - \delta^2/2}{z_j} & \text{if} \quad \rho_j > \delta^2/2 \end{cases}$$
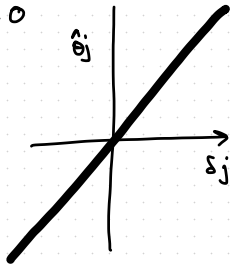
LASSO (SOFT) THRESHOLDING

$$\hat{\theta}_j = \begin{cases} \dfrac{\rho_j + \delta^2/2}{z_j} & \text{if} \quad \rho_j < -\delta^2/2 \\[2mm] 0 & \text{if} \quad -\dfrac{\delta^2}{2} \leq \rho_j \leq \dfrac{\delta^2}{2} \\[2mm] \dfrac{\rho_j - \delta^2/2}{z_j} & \text{if} \quad \rho_j > \delta^2/2 \end{cases}$$

LASSO (SOFT) THRESHOLDING

$$\hat{\theta}_j = \begin{cases} \dfrac{\rho_j + \delta^2/2}{z_j} & \text{if} \quad \rho_j < -\delta^2/2 \\[2mm] 0 & \text{if} \quad -\dfrac{\delta^2}{2} \leq \rho_j \leq \dfrac{\delta^2}{2} \\[2mm] \dfrac{\rho_j - \delta^2/2}{z_j} & \text{if} \quad \rho_j > \dfrac{\delta^2}{2} \end{cases}$$

LASSO (SOFT) THRESHOLDING

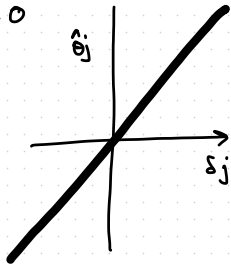$\delta^2 = 0$

$\hat{\theta}j$
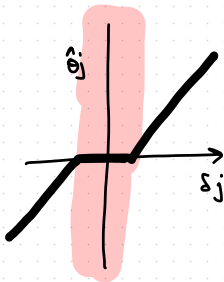
δj

NO REGULARISATION

LASSO (SOFT) THRESHOLDING

$\delta^2 = 0$

NO   REGULARISATION

REGULARISATION
⇓
SPARSITY